



Honors Theses at the University of Iowa

Fall 2019

Releasing the Fantasy: Wittgenstein's Critique of Consciousness

Killian Beck

Follow this and additional works at: https://ir.uiowa.edu/honors_theses



Part of the [History of Philosophy Commons](#), [Philosophy of Language Commons](#), and the [Philosophy of Mind Commons](#)

This honors thesis is available at Iowa Research Online: https://ir.uiowa.edu/honors_theses/318

RELEASING THE FANTASY: WITTGENSTEIN'S CRITIQUE OF CONSCIOUSNESS

by

Killian Beck

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Philosophy

David Stern
Thesis Mentor

Fall 2019

All requirements for graduation with Honors in the
Philosophy have been completed.

Richard Fumerton
Philosophy Honors Advisor

Releasing the Fantasy: Wittgenstein's Critique of Consciousness

Killian Beck

Introduction

My work defends Wittgenstein's continued relevance to philosophy of mind by presenting a close exegesis of §420 of the *Philosophical Investigations*, a remark in which he anticipates contemporary debates concerning the conceivability of so-called "zombies", or imaginary creatures who lack consciousness, but are otherwise identical to human beings. In §1, I survey some of the major historical developments that led to the emergence of the idea of zombies in the mid-1970's, before discussing David Chalmers' use of the idea in a modal argument against physicalism. In §2, I turn to the work of one of Chalmers' most prominent opponents, Daniel Dennett, whose rejection of the conceivability of zombies is informed by the scientifically-minded approach to consciousness that he advocates. Despite avowing his influence, I argue in §§3 and 4 that Dennett's externalist approach to consciousness diverges sharply from that taken by Wittgenstein.

1.

In his paper, “Wittgenstein, Wittgensteinianism, and the Contemporary Philosophy of Mind – Continuities and Changes”, Ansgar Beckermann accounts for the emergence of what he refers to as a “new, post-Wittgensteinian orthodoxy” in contemporary philosophy of mind.¹ According to Beckermann, during the first few decades following the posthumous publication of the *Philosophical Investigations*, “Wittgenstein was the measure of all things in philosophy and especially in the philosophy of mind.”² The Wittgensteinian orthodoxy of the 1950’s and 60’s drew inspiration from the book’s private language discussion and Wittgenstein’s remarks on the use of psychological terms, more generally. As Beckermann explains:

The predominant view in the 50s and 60s was a view that one could call the “criteriological account.” According to the proponents of this view Wittgenstein has shown by means of considerations on the meaning of linguistic expressions in general that there can be no mental states without behavioural criteria. Pain behaviour is not just a symptom of the mental state pain, but a criterion. That is to say, pain behaviour is corrigible evidence that somebody is in pain, but for semantic reasons it is, in a certain way, also sufficient evidence. For semantic reasons, it is true that if a person shows this behaviour and there is no evidence to the contrary, then this person is in pain.³

One of the most important implications for this view was its apparent dissolution of the mind-body problem. For on the “criteriological account” of mind, the distinction between the mental and the physical is not understood as a difference between the types of phenomena, to which, concepts of the former and the latter correspond, respectively. It is, rather, a failure to notice the

¹ Beckermann (2004), p. 287.

² Ibid.

³ Ibid., p. 288.

difference between our use of mental and physical concepts that creates the illusion of a substantial difference between mental and physical phenomena, respectively.

The Wittgensteinian orthodoxy in philosophy of mind began to wane after the causal character of mental explanation came under dispute in the late 1960's.⁴ Broadly speaking, proponents of criteriological accounts of mind would deny that statements involving mental concepts are explained by underlying mental processes or events, but are accounted for in terms of patterns of behaviors, instead. This view came under pressure by causalists who believed that mental explanations do in fact exhibit a causal character and should therefore be analyzed as such. They believed that criteriological accounts had gone too far in denying that an agent's cognitive states – her beliefs, desires, etc. – are causally related to her behavior. In their view, for example, the concept, “pain” in the utterance, “She is holding her cheek because she is in pain”, is more naturally analyzed as an explanation for her pain behavior than as a reference to a certain pattern of behavior. By treating mental concepts as “theoretical concepts”, causalists believed they could preserve the causal character expressed by statements involving mental explanations, without taking those terms to refer to inner, private mental entities.⁵

Beckermann thus identifies the dispute over the causal character of mental concepts at the end of the 1960's and into the early 1970's as an important turning point in Wittgenstein's reception in philosophy of mind – a dispute in which the causalists, according to Beckermann, “clearly won the day.”⁶ Although Beckermann welcomes the new, post-Wittgensteinian orthodoxy in the philosophy of mind, he does regret one of its consequences: the revival of the mind-body problem.

⁴ Ibid., p. 305.

⁵ Ibid., p. 301.

⁶ Ibid., p. 305.

[I]n the mid-70s something began to happen within the framework of the new orthodoxy which could well be regarded as a return to Cartesianism. The starting point was Thomas Nagel's seminal paper "What is it like to be a bat?". The considerations of Nagel, Jackson, Levine, Chalmers and many others all seem to point to the same result, namely, that at least phenomenal states have characteristic features that in the last consequence are not public, since they are neither tied to typical behaviours nor to causal roles. The idea of the philosophical zombie was born – the idea of a being that in all situations says exactly the same as I say, and does exactly the same as I do, but whose phenomenal states are – on this assumption – either connected with radically different qualia or with none at all.⁷

Among those figures Beckermann lists, David Chalmers has made especially far-reaching use of philosophical zombies. His "zombie argument", as I'll henceforth refer to it, is a modal argument against materialism. It depends, foremost, on a distinction he draws between two concepts of mind, which he terms the "psychological" and "phenomenal" concepts of mind, respectively.⁸ Chalmers describes the former as "the concept of mind as the causal or explanatory basis for behavior. A mental state is conscious in this sense if it plays the right sort of causal role in the production of behavior..."⁹ On the latter concept, "mind is characterized by the way it *feels*"¹⁰, i.e. in terms of the "*something it is like*"¹¹, or *qualia* of experience. Thus, a zombie, for Chalmers, is a creature for whom the "phenomenal" concept of mind necessarily could not apply:

What is going on in my zombie twin? He is physically identical to me, and we may as well suppose that he is embedded in an identical environment. He will certainly be identical to me

⁷ Ibid., p. 305.

⁸ Chalmers (1996), p. 10.

⁹ Ibid.

¹⁰ Ibid.

¹¹ Ibid., p. 3.

functionally: he will be processing the same sort of information, reacting in a similar way to inputs, with his internal configurations being modified appropriately and with indistinguishable behavior resulting. He will be *psychologically* identical to me, in the sense developed in Chapter 1. He will be perceiving the trees outside, in the functional sense, and tasting the chocolate, in the psychological sense. All of this follows logically from the fact that he is physically identical to me, by virtue of the functional analyses of psychological notions. He will even be “conscious” in the functional senses described earlier—he will be awake, able to report the contents of his internal states, able to focus attention in various places, and so on. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie.¹²

Chalmers takes the conceivability of zombies to be entirely unproblematic, and thus readily employs the idea in a modal argument against physicalism:

- (1.) It is conceivable that there be zombies.
- (2.) If it is conceivable that there be zombies, it is metaphysically possible that there be zombies.
- (3.) If it is metaphysically possible that there be zombies, then consciousness is nonphysical.
- (4.) Consciousness is nonphysical.¹³

Chalmers contrasts “metaphysically possible” with “naturally possible”, and thus admits that zombies “probably cannot exist in our world, with its laws of nature.”¹⁴ Nonetheless, if something is metaphysically possible, then it could have existed, or the universe could have been such that it existed. Thus, as Chalmers argues, “if there is a metaphysically possible universe that is physically identical to ours but that lacks consciousness, then consciousness must be a further, nonphysical component of our universe.”¹⁵ In other words, he believes that consciousness isn’t

¹² Ibid., pp. 84-5.

¹³ Chalmers (2002), pp. 247-72.

¹⁴ Ibid., p. 249.

¹⁵ Ibid.

entailed by physical facts, alone, since the metaphysical possibility of zombies implies that the very same set of facts would be consistent with the non-existence of consciousness. Thus, Chalmers concludes, there exists an “*explanatory gap* between the physical level and conscious experiences.”¹⁶ What Chalmers refers to as the “hard problem” of consciousness, then, is precisely the problem of bridging that gap.¹⁷

Although not entirely dismissive of those who deny the conceivability of zombies (or “Type-A materialists” as he refers to them), Chalmers does admit that the question of whether zombies are conceivable is ultimately a matter of intuition.¹⁸ Whether one shares Chalmers’ intuition concerning the conceivability of zombies will likely depend on whether one approves his distinction between the “psychological” and “phenomenal” concepts of mind, viz. that the latter characterizes mind in terms of the qualia that (to use the word Chalmers favors) “accompany” experience. More broadly, Chalmers’ intuition would likely appeal to those who believe that conscious experience has an ineradicably subjective character – a conviction one might express by insisting that one’s experience of the world is from a *particular point of view*, so to speak.¹⁹ In any case, Chalmers seems to take some comfort in his observation that “the intuition appears to be shared by the large majority of philosophers, scientists and others; and it is so strong that to deny it, a type-A materialist needs exceptionally powerful arguments. The result is that even among materialists, type-A materialists are a distinct minority.”²⁰

2.

¹⁶ Chalmers (1996), p. 94.

¹⁷ Chalmers (1995).

¹⁸ Chalmers (1996), p. 85. See also: Chalmers (2002), p. 253.

¹⁹ Cf. Nagel (1974).

²⁰ Chalmers (2002), p. 253.

One particularly loud voice in that minority is Daniel Dennett, whose criticism of the idea and its supporters often verges on the polemical:

[W]hen philosophers claim that zombies are conceivable, they invariably underestimate the task of conception (or imagination), and end up imagining something that violates their own definition. This conceals from them the fact that the philosophical concept of a zombie is sillier than they have noticed.²¹

Central to Dennett's hardline functionalist account of mind is his denial of the putative "explanatory gap" between the physical and the phenomenal, for he regards the latter category as entirely vacuous; there simply are no "remnants" of experience that would be left unexplained by a complete account of those phenomena that, say, Chalmers takes the "psychological" concept of mind to designate.²² In attempting to distinguish between the "psychological" and "phenomenal" concepts of mind, respectively, Chalmers commits what Dennett calls the "fallacy of subtraction", which he illustrates accordingly:

Supposing that by an act of stipulative imagination you can remove consciousness while leaving all cognitive systems intact - a quite standard but entirely bogus feat of imagination - is like supposing that by an act of stipulative imagination, you can remove health while leaving all bodily functions and powers intact. If you think you can imagine this, it's only because you are confusedly imagining some health-module that might or might not be present in a body. Health isn't that sort of thing, and neither is consciousness.²³

Dennett's opposition to those who maintain the conceivability of zombies amounts to a denial that there could be anything that a zombie could be imagined to lack, if it's assumed that zombies are functionally identical to humans. For Dennett, then, the question of whether zombies

²¹ Dennett (1995), p. 322.

²² Dennett (2002).

²³ Ibid., p. 325.

are conceivable has a straightforward answer: no.

On Beckermann's assessment of the new orthodoxy in philosophy of mind, "those who claim that philosophical zombies are possible have strayed a step too far from Wittgensteinianism."²⁴ It is, perhaps, a sign of Dennett's Wittgensteinianism, then, that he represents an exception to the new orthodoxy in this regard. Indeed, Dennett considers Wittgenstein among the most formative of his philosophical influences and even characterizes his own work as "a kind of redoing of Wittgenstein's attack on the objects of conscious experience."²⁵ According to Dennett's own self-assessment, then, his work can be viewed as a proxy through which Wittgenstein's legacy continues to shape the philosophy of mind.

Against this, I argue that Dennett's approach to questions about consciousness diverges from Wittgenstein's in a number of fundamental respects, such that the overall character of the former's views is, at most, only distantly "Wittgensteinian." To this end, I begin by considering how Dennett conceives his relation to Wittgenstein before discussing his "heterophenomenological" approach to mind. I argue that Dennett's misreading of the *Investigations* leads him to overestimate his proximity to Wittgenstein, the consequences of which are further exhibited by the decidedly *un*-Wittgensteinian character of his heterophenomenological method.

On the status of qualia, for instance, Dennett regards Wittgenstein as an ally.²⁶ Citing his "beetle in a box" story, Dennett claims that Wittgenstein, likewise, denied the existence of qualia, before specifying the finer differences in their views:

I choose to take what may well be a more radical stand than Wittgenstein's. Qualia are not even

²⁴ Beckermann (2004), p. 305.

²⁵ Dennett (1991), p. 22.

²⁶ Dennett (2002), p. 230.

“something about which nothing can be said”; “qualia” is a philosopher’s term which fosters nothing but confusion, and refers in the end to no properties or features at all.²⁷

Although it isn’t entirely clear from the above passage, Dennett apparently suggests that to regard qualia as “something about which nothing can be said” would render them ineffable – a stance that would thus fall short of what his own eliminative materialism prescribes, i.e., “deny[ing] resolutely the existence of something real or significant”²⁸ – in this case, “qualia.”

In any event, Dennett presents a distorted view of his relation to Wittgenstein by conceiving of it merely in terms of who between them took the “more radical stand”. On the contrary, far from supporting his assessment, the remark that Dennett cites from the *Investigations* suggests a rather more complicated situation. Quoted in full, §304 reads:

304. “But you surely admit that there is a difference between pain behavior with pain and pain-behavior without pain.” — Admit it? What greater difference could there be? — “And yet you again and again reach the conclusion that the sensation itself is a Nothing.” — Not at all. It’s not a Something, but not a Nothing either! The conclusion was only that a Nothing would render the same service as a Something about which nothing could be said. We’ve only rejected the grammar which tends to force itself on us here.

The paradox disappears only if we make a radical break with the idea that language always functions in one way, always serves the same purpose: to convey thoughts – which may be about houses, pains, good and evil, or whatever.²⁹

By denying that “qualia” refers to any “properties or features at all”, Dennett apparently ignores Wittgenstein’s warning that, “a Nothing would render the same service as a Something about which nothing could be said.” From the perspective of §304, then, Dennett’s view is no better off

²⁷ Ibid. See also: Stern (2007).

²⁸ Ibid., p. 226.

²⁹ Wittgenstein (2009). Hereafter, all references to the *Investigations* will consist of in-line citations.

than that which it opposes. For by conceiving the problem as that of determining whether “qualia” refers to anything at all, Dennett presumes an oversimplified account of how our mental concepts work – viz., on the broadly referentialist “model of ‘object and name’” (§293). This presumption is hardly an isolated feature of Dennett’s project, but underwrites his entire approach to consciousness – a point that I hope to make clear shortly.

Dennett advertises his heterophenomenological approach to consciousness as a third-person, scientifically-minded replacement for the introspective methods of traditional phenomenology.³⁰ Accordingly, Dennett rejects that a subject occupies an authoritative position with respect to her own consciousness. Instead, Dennett’s heterophenomenological approach involves an investigator who – in an anthropological fashion – considers a subject’s self-reports against the background of other relevant information about her and her immediate environment. The role of the investigator, Dennett explains:

[...] is to compose a catalogue of *what the subject believes to be true about his or her conscious experience*. This catalogue of beliefs fleshes out the subject’s *heterophenomenological world*, the world according to S — the subjective world of one subject — not to be confused with the real world. The total set of details of heterophenomenology, plus all the data we can gather about concurrent events in the brains of subjects and in the surrounding environment, comprise the total data set for a theory of human consciousness. It leaves out no objective phenomena and no subjective phenomena of consciousness.³¹

The investigator, moreover, is to adopt what Dennett calls an “intentional stance” towards the subject, which consists in remaining maximally neutral with regard to questions concerning the

³⁰ Dennett (2003), p. 19.

³¹ Ibid., p. 20.

content of the subject's self-reports.³² Moreover, the intentional stance compels the investigator to remain neutral on the question of whether the subject of investigation is even conscious at all, presupposing only that the subject is an "intentional system, capable of meaningful communication."³³ The "primary", or "*raw*, uninterpreted" data available to the investigator are the subject's utterances.³⁴ The investigator then interprets the subject's utterances as "'verbal judgements' expressing her beliefs", before, in the end, interpreting those judgements to involve the subject's beliefs about her own (conscious) experience.³⁵

The heterophenomenological worlds generated by this process are not taken to correspond with, or describe, the subject's conscious experience, itself, but merely the subject's beliefs about her experiences, where such beliefs are understood to describe how the subject's conscious experiences *seem* to her, as opposed to how they *really* are.³⁶ Dennett clarifies that such beliefs are to be treated "from the *intentional stance* as *theorists' fictions* similar to centres of mass, the Equator, and parallelograms of force."³⁷ The upshot, as Dennett explains, is that, "Heterophenomenology is the beginning of a science of consciousness, not the end. It is the organization of the data, a catalogue of *what must be explained*, not itself an explanation or a theory."³⁸

Although Dennett intends heterophenomenology as a maximally neutral data-gathering method, it's accompanied by a peculiarly dogmatic outlook that admits neither the diversity of

³² Ibid.

³³ Ibid. p. 21. The feasibility of Dennett's intentional stance is taken up by Cerbone (2019), which he contrasts with Wittgenstein's notion of an "attitude towards a soul" (PPF §22).

³⁴ Ibid., p. 21.

³⁵ Ibid.

³⁶ Ibid., p. 20.

³⁷ Ibid.

³⁸ Ibid., p. 28.

language nor indefiniteness of conscious experience. To start, the three-stage interpretative procedure by which a subject's utterances are ultimately interpreted as truth-functional beliefs depends on the un-argued assumption that the variety of a subject's speech acts – of which Dennett lists “reporting, questioning, correcting, requesting” – exhibit the same logical form.³⁹ That is, Dennett alternately refers to the interpretation of the subject's utterances as “speech acts” or “judgements expressing those beliefs”, and thus implicitly takes the former to be analyzable as judgements with truth-functional contents. At least, he doesn't take it to be a distinct interpretative move for the heterophenomenological investigator to treat her subject's utterances as speech acts or as judgements with truth-functional contents, and vice versa.

In any case, even if it's supposed that the heterophenomenological investigator is concerned only with those of the subject's utterances that can be analyzed as truth-functional judgements – her “reporting” speech acts, say – Dennett doesn't account for the possibility of their successful identification, nor does he explain how such reports are supposed to be interpreted.⁴⁰ Indeed, Dennett takes such matters of interpretation to be entirely unproblematic, and claims that the investigator's “task of unifying the interpretation of all the verbal judgements into a heterophenomenological world is akin to reading a novel.”⁴¹

Dennett's analogy not only suggests his impatience for the sorts of considerations I've raised, but moreover reveals his commitment to the very conception of language from which Wittgenstein urged a “radical break” (§304), i.e., “the idea that language always functions in one way, always serves the same purpose: to convey thoughts which may be about houses, pains, good and evil, or whatever” (§304). Accordingly, in taking the heterophenomenological subject's

³⁹ Ibid., p. 20.

⁴⁰ A similar concern is raised by Hutto (1995).

⁴¹ Dennett (2003), p. 22.

utterances to express truth-functional contents about her experience, Dennett considers only one of two alternatives. Either:

- (i.) the subject's utterance expresses a true belief about her experience, in which case the subject's description of her internal state is theoretically confirmable; or,
- (ii.) the belief expressed by the subject's report about her experience is false, in which case the subject's description of her internal state is not theoretically confirmable.

Those of a subject's reports that fail to yield theoretically confirmable descriptions are "demoted", as it were, to talk of fictitious objects. Dennett thereby uses "fiction" to invoke an ontological contrast with reality – viz., to distinguish between that which correctly and incorrectly describes reality – a distinction that expedites his treatment of philosophical difficulties by, for example, allowing the easy disposal of such philosophically problematic ideas as "qualia" or "zombies."

As a term of philosophical criticism, then, the sense in which Dennett uses "fiction" is entirely distinct from that in which Wittgenstein uses it:

307. "Aren't you nevertheless a behaviorist in disguise? Aren't you basically saying that everything except human behavior is a fiction?" – If I speak of a fiction, then it is of a *grammatical* fiction (§307).

A "grammatical fiction" isn't, for Wittgenstein, a *mistaken* idea of the way a word functions – supported, as it were, by a metaphysical distinction between correct and incorrect uses of language, respectively – but rather, as Wittgenstein puts it, a mere "picture" that "stands in the way of our seeing the use of a word as it is" (§305). Wittgenstein's goal isn't to avoid the difficulties entailed by such pictures by claiming that they fail to accurately describe reality, but by targeting their sources in the inclinations that, while philosophizing, lead to our "bewitchment" by them (§109). In the next section, I elaborate on this aspect of Wittgenstein's

philosophical outlook by considering his treatment of philosophical difficulties concerning consciousness in §§412-27 of the *Investigations*. I focus, in particular, on §420 – a remark in which Wittgenstein can be seen to anticipate contemporary debates concerning the conceivability of zombies.

3.

In §420 of the *Investigations*, Wittgenstein considers a nearly identical question to that of whether zombies are conceivable:

420. But can't I imagine that people around me are automata, lack consciousness, even though they behave in the same way as usual? – If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a little uncanny. But just try to hang on to this idea in the midst of your ordinary intercourse with others – in the street, say! Say to yourself, for example: "The children over there are mere automata; all their liveliness is mere automatism." And you will either find these words becoming quite empty; or you will produce in yourself some kind of uncanny feeling, or something of the sort.

Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another; the cross-pieces of a window as a swastika, for example.

Before moving on, I want to clarify, briefly, the most obvious difference between the notions of a zombie and an automaton. While an automaton is taken to lack consciousness altogether, a zombie is more narrowly defined as lacking, merely, *phenomenal* consciousness. Recalling Chalmers' use of the notion, a zombie is granted the full range of conscious experiences enjoyed by its conscious human counterparts, and is deprived only of the *what-it's-like* or "qualia" of conscious experience. Although this difference might seem substantial, I believe it can be set aside without too much cost for the purposes of this essay.

Unlike Dennett, who flatly denies the conceivability of zombies, one doesn't find a similarly straightforward answer to the interlocutory voice's question in §420. On the contrary, Wittgenstein's narrator at first seems to concede that we could imagine that others were automata

(at least in solitude) only to intimate that we could not (at least not without difficulty) hold onto such an idea amidst our ordinary intercourse with others.⁴²

However, that Wittgenstein's narrator appears to equivocate depends on our presuming that the interlocutory voice's question is itself intelligible. Doing so would not only foreclose the possibility of our appreciating the remark's subtler argumentative structure, but would indicate a more general failure to acknowledge one of the defining characteristics of Wittgenstein's outlook in the *Investigations*:

374. The great difficulty here is not to present the matter as if there were something one *couldn't* do. As if there really were an object, from which I extract a description, which I am not in a position to show anyone. — And the best that I can propose is that we yield to the temptation to use this picture, but then investigate what the *application* of the picture looks like.

This outlook appears most prominently in the private language discussion of §§243ff., in which the following question (similar to that with which §420 begins) is addressed: “But is it conceivable that there be a language in which a person could write down or give voice to his inner experiences – his feelings, moods, and so on – for his own use?” (§243), followed by the qualification that “The words of this language are to refer to what only the speaker can know – to his immediate private sensations. So another person cannot understand the language” (§243). As the discussion that follows §243 indicates, Wittgenstein's concern is not to show the

⁴² My use of “interlocutory voice” and “Wittgenstein's narrator” to distinguish the dialogic participants of §420 is a device I've borrowed from Stern (2004), whose approach to the *Investigations* emphasizes its “profoundly dialogical character” (p. 37). In addition to these voices, he identifies a third voice which “provides an ironic commentary” on the exchanges between the other two, consisting “partly of objections to assumptions the debaters take for granted, and partly of platitudes about language and everyday life they have both overlooked” (p. 22). Although he takes the commentator to come closest to expressing the author's views, Stern ultimately warns against identifying any of these voices with that of that author's. On this aspect of the text's style and its methodological role, see Stern (2017). See also Cavell (1962), which Stern credits as the first reading of the *Investigations* to recognize its multiplicity of voices.

impossibility of such a (private) language, but rather, as David Stern argues, to lead us to see that the “very idea of a private language is illusory, and falls apart upon closer examination.”⁴³

In a similar vein, Cavell denies that the upshot of Wittgenstein’s private language discussion concerns a “failure of imagination” or the “non-existence of a private language”, but rather that “there is nothing of the sort to imagine, or rather that when we as it were try to imagine this we are imagining something other than we think.”⁴⁴ Cavell echoes the language Wittgenstein employs in §374 – that of *yielding to temptation* – in the way he characterizes the “tone” of the question raised in §243 as that of “someone allowing a fantasy to be voiced”, such that Wittgenstein seeks not the denial of this fantasy, but its *release*.⁴⁵

I believe that §420 is occupied by a similar concern, i.e., that of releasing the fantasy expressed in the interlocutory voice’s question, “But can’t I imagine that other people are automata, lack consciousness, even though they behave in the same way as usual?” My goal is to show that careful attention to the remarks that precede §420 should make us suspicious of the interlocutory voice’s question, i.e., that it should hardly come as a surprise that the interlocutory voice’s question is *not* intelligible, nor treated as such, given the set-up in §§412-19.

That the interlocutory voice’s question in §420 begins with the subordinating conjunction, “But”, suggests that it takes the form of a response to, or continuation of, a broader discussion. Indeed, §§412-27 constitute an extended discussion on the topic of consciousness, a discussion precipitated by the question with which §412 opens: “The feeling of an unbridgeable gulf between consciousness and brain processes: how come that this plays no role in reflections of ordinary life?” This question contains, in germ, Chalmer’s “hard problem” of consciousness

⁴³ Stern (2011).

⁴⁴ Cavell (1979), p. 344.

⁴⁵ Ibid.

which I discussed in Section 1, i.e., that of accounting for the putative explanatory gap between the physical and the phenomenal. Rather than seeking an account of consciousness that bridges that gap, however, Wittgenstein shifts the focus onto the ways in which *we* create the illusion of one:

The idea of a difference in kind is accompanied by slight giddiness – which occurs when we are doing logical tricks. (The same giddiness attacks us when dealing with certain theorems in set theory.) When does this feeling occur in the present case? It is when I, for example, turn my attention in a particular way on to my own consciousness and, astonished, say to myself: “THIS is supposed to be produced by a process in the brain!” – as it were clutching my forehead. – But what can it mean to speak of “turning my attention on to my own consciousness”? There is surely nothing more extraordinary than that there should be any such thing! What I described with these words (which are not used in this way in ordinary life) was an act of gazing. I gazed fixedly in front of me – but not at any particular point or object. My eyes were wide open, brows not contracted (as they mostly are when I am interested in a particular object). No such interest preceded this gazing. My glance was vacant; or again, *like* that of someone admiring the illumination of the sky and drinking in the light.

Note that the sentence which I uttered as a paradox (“THIS is produced by a brain process!”) has nothing paradoxical about it. I could have said it in the course of an experiment whose purpose was to show that an effect of light which I see is produced by stimulation of a particular part of the brain. – But I did not utter the sentence in the surroundings in which it would have an everyday and unparadoxical sense. And my attention was not such as would have been in keeping with that experiment. (If it had been, my gaze would have been intent, not vacant) (§412).

Under criticism is the tendency to take for granted the sense of an expression without first considering the sorts of occasions on which – and the associated activities in which – it would

make sense to use it – i.e., the various *language-games* in which an expression is meaningfully used. The inclination to consider the sense of an expression apart from its use is associated with a particular conception of language according to which the meaning of a word can be given independently of that word's employment on any particular occasion, i.e., that the context and circumstances in which a word is used are incidental to its meaning – as if, according to Wittgenstein, “the meaning were an aura the word brings along with it and retains in every kind of use” (§117).⁴⁶ Against this, Wittgenstein insists that language is meaningful only in light of such features, such that determining the sense of an expression must involve a consideration of concrete circumstances in which it's used. For Wittgenstein, then, the task of imagining various language-games in which an expression is meaningfully used is to consider the correspondingly distinct possible senses for that expression. Thus, as James Conant explains, determining the sense of an expression “is a matter of perceiving – of the various possible contributions which circumstances of use might make – what sort of contributions the actual circumstances are most reasonably taken to make.”⁴⁷

The point of such an exercise is not, however, to call to mind, say, the determinate rules for the meaningful use of an expression; Wittgenstein was keen to avoid any impression that he was interested in systematizing language in such a way (cf. §§130-3). Rather, as Wittgenstein clarifies, the exercise is itself occasioned by cases of philosophical confusion in which “language

⁴⁶ Cf. §120: “People say: it's not the word that counts, but its meaning thinking of the meaning as a thing of the same thing as the word, even though different from the word. Here the word, there the meaning. The money, and the cow one can buy with it. (On the other hand, however: money, and what can be done with it.)”

⁴⁷ Conant (2005), p. 61.

is, as it were, idling, not when it is doing work” (§132).⁴⁸ Cavell helpfully characterizes such cases as follows:

[W]e are led to speak ‘outside language games’, consider expressions apart from, and in opposition to, the natural forms of life which give those expressions the force they have. [...] What is left out of an expression if it is used “outside its ordinary language game” is not necessarily what the *words* mean (they may mean what they always did, what a good dictionary says they mean), but what we mean in using them when and where we do. The point of saying them is lost.⁴⁹

Cavell highlights an important dimension of Wittgenstein’s anti-metaphysical conception of philosophical problems by emphasizing the internal relation between philosophical confusion and the emptiness of what *we* say – *our* failure to make sense (cf. §§111, 123, and 125). This failure, according to Conant, “is to be traced to a failure on the part of the speaker to project that string of words into a new context in a fashion which admits of a stable reading – in a fashion which admits of our being able to perceive in the sentence, when we view it *against the background* of its circumstances of use, a coherent physiognomy of meaning.”⁵⁰ We are, in such cases, merely under the *illusion* of meaning something.

In §§412-27, Wittgenstein identifies the source of our failure to make sense in a misleading picture of consciousness as essentially private, according to which there exists an asymmetry between the (type of) access I have to my own consciousness, and the (type of) access I have to that of another’s, i.e., that I’m *directly*, or non-inferentially acquainted with my own consciousness, but only *indirectly*, or inferentially acquainted with that of another’s. Among

⁴⁸ Cf. §38: “For philosophical problems arise when language *goes on holiday*.”

⁴⁹ Cavell (1979), p. 207; my emphasis.

⁵⁰ Conant (2005), p. 61.

the difficulties precipitated by this picture, then, is that of accounting for the generality of the concept of consciousness, viz. the criteria by which consciousness is ascribed in particular cases. For if consciousness is essentially private, what justifies my belief that others are conscious?

In addressing this difficulty, Wittgenstein, unlike Dennett, doesn't reject that consciousness is private and defend its antithesis – i.e., that consciousness is, essentially, public. Rather, Wittgenstein is concerned to show that we simply can't make sense of the idea that consciousness is essentially private by considering a series of cases in which we're merely under the illusion of meaning something definite by it – cases in which, to borrow Cavell's expression, we "mean something other than [we] think."⁵¹ Accordingly, the dialogic exchanges between Wittgenstein's narrator and an interlocutory voice in §§416-20 can be read as unsuccessful attempts to make sense of the idea of an asymmetry between first and third-personal statements about consciousness. §§416-7 addresses the idea that (i) first-personal uses of the concept of consciousness involve direct, non-inferential judgements while §§418-20 address the latter part of this idea, namely, that (ii) third-personal uses of the concept of consciousness involve indirect, inferential judgements.

In §416, an interlocutory voice suggests that ascriptions of consciousness are conditioned by more explicit forms of agreement among human beings: "Human beings agree in saying that they see, hear, feel, and so on (even though some are blind and some are deaf). So they are their own witnesses that they have *consciousness*." The idea would seem to be that remarks such as, "I see...", "I hear...", "I feel..." etc., are reports of consciousness. Furthermore, that distinct individuals use these expressions would therefore seem to imply the existence of something in common between them, with which they're directly acquainted, i.e., *consciousness*.

⁵¹ Ibid., p. 34; my emphasis.

Wittgenstein draws the reader's suspicion to this claim with the inclusion of the parenthetical "(even though some are blind and some are deaf)" (§416), since the line of reasoning employed by the interlocutory voice would apparently fail to account for individuals lacking such capacities as seeing or hearing. Nevertheless, the full weight of Wittgenstein's criticism is borne through his narrator's exclamation at the idea's utter absurdity:

But how strange this is! Whom do I really inform if I say "I have consciousness"? What is the purpose of saying this to myself, and how can another person understand me? – Now, sentences like "I see", "I hear", "I am conscious" really have their uses. I tell a doctor "Now I can hear with this ear again", or I tell someone who believes I am in a faint "I am conscious again", and so on (§416).

Wittgenstein's narrator criticizes the interlocutory voice's assumption that the remark "I have consciousness" takes the form of a report by contrasting it with remarks such as "I see", "I hear", and "I am conscious" (§416). While we can readily imagine cases in which those latter remarks could be used to inform another of something, it's difficult to imagine cases in which the former could be used informatively. The suggestion, then, is that because of its distinct grammar, first-personal uses of the concept of consciousness can't be taken to involve matter-of-factual judgements in the same sense in which perceptual reports do. Unlike perceptual reports, first-personal uses of the concept of consciousness are informative in quite particular circumstances, such as those in which, to use the example from §416, we tell another person, "I am consciousness" after having fainted.

In the following remark, Wittgenstein's narrator responds to the interlocutory voice's suggestion in §416 that human beings are "their own witnesses that they have *consciousness*":

“Do I observe myself then, and perceive that I am seeing or conscious? And why talk about observation at all? Why not simply say “I perceive I am conscious”? – But what are these words ‘I perceive’ for here – why not say ‘I am conscious’?” (§417).

In response, an interlocutory voice suggests, “But don’t the words ‘I perceive’ here show that I am attending to my consciousness?” (§417). This suggestion, however, won’t do, for as Wittgenstein’s narrator subsequently points out, the non-superfluous addition of “I perceive” implies that the remark “I perceive I am conscious” will no longer mean “I am conscious”, but rather, “that my attention is focused in such-and-such a way” (§417).

The difficulty I noted earlier regarding the generality of the concept of consciousness comes to the fore in §418:

418. Is my having consciousness a fact of experience? –

But doesn’t one say that human beings have consciousness, and that trees or stones do not? – What would it be like if it were otherwise? – Would human beings all be unconscious? – No; not in the ordinary sense of the word. But I, for instance, would not have consciousness — as I now in fact have it.

Joachim Schulte helpfully identifies two (not altogether separable) strands of thought in the interlocutory voice’s question in §418: “[1] is it an empirical (as opposed, say, to a logical) fact that [2] I (as opposed to a different sort of creature) have consciousness?”⁵² Wittgenstein’s narrator addresses these questions by asking us to consider what it would be like if things were otherwise, i.e., what it would be like if it weren’t the case that human beings (as opposed to trees and stones) had consciousness.

⁵² Schulte (2017), p. 306.

The view that emerges from §§418-9 is that ascriptions of consciousness in particular cases depend very little on extra-grammatical criteria, in contrast to those cases in which, for example, a fictional anthropologist might determine whether a tribe has a chief:

419. In what circumstances shall I say that a tribe has a *chief*? And the chief must surely have *consciousness*. Surely he mustn't be without consciousness!

The above remark becomes clearer if we consider it alongside a related scenario that Wittgenstein discusses in *Remarks on the Foundations of Mathematics*:

We come to an alien tribe whose language we do not understand. Under what circumstances shall we say that they have a chief? What will occasion us to say that this man is the chief even if he is more poorly clad than others? The one whom the others obey—is he without question the chief?

What is the difference between inferring wrong and not inferring? [...] Consider this.⁵³

We can imagine, therefore, various criteria by which a fictional anthropologist would infer whether a tribe has a chief, for instance, by determining whether one member seems to be giving the other orders. Of course, it's possible not only that we might fail to identify which member of the tribe is the chief, but that we're moreover wrong in presuming that the tribe even *has* a chief. This latter possibility effectively illustrates the dis-analogy between determining, on the one hand, whether a tribe has a chief, and whether, on the other hand, the chief has consciousness. While in the former case, there's a clear sense in which we might infer wrongly, and therefore a clear "difference between inferring wrong and not inferring" in the latter case, there isn't a similarly clear difference between inferring wrong and not inferring, for it isn't clear what it would mean to be wrong in assuming that a given member of the tribe is conscious.

⁵³ Wittgenstein (1983), p. 352; my emphasis.

The dis-analogy in §419 can therefore be read as drawing our suspicion to the idea that ascriptions of consciousness involve hypothetical judgments, or opinions – that such judgements concern empirical facts, or “facts of experience” (§418). Of course, human beings in general can be said to “have consciousness,” but this amounts to a rather trivial claim, one that would properly be said, as Wittgenstein puts it in a subsequent passage of *RFM*, to belong to the “form of the natural-historical proposition” (p. 353), or the form of those “remarks concerning the natural history of human beings; not curiosities, however, but facts that no one has doubted, which have escaped notice only because they are always before our eyes” (§415). In other words, that human beings “have consciousness” is not an inference we make on the basis of observable criteria. Rather, it is a *grammatical remark*, or a remark about the grammar of the concept of ‘consciousness.’ Consider, in this connection, the following: “[...] only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious” (§282). Relatedly, to ask whether a given non-human animal has consciousness is not to ask whether it possesses something that human beings, likewise, possess – whether, in other words, we’re to include it in the fantasy of ourselves. On the contrary, to ask whether it has consciousness is to admit uncertainty about our relation to it. This implies not only that we might have an insufficient understanding of it, but that we might also have an insufficient understanding of ourselves.

We can therefore see, on the basis of the remarks that precede §420, that the interlocutory voice is, in a sense, *forced* to the assumption that one could imagine that others were automata. The interlocutory voice wants to maintain, despite what Wittgenstein’s narrator suggests in §§416-19, that having consciousness, or being conscious, is an empirical fact, or a “fact of experience” (§418). Accordingly, the interlocutory voice must admit the possibility that human

beings lacked consciousness per the terms of the counterfactual in §418 – i.e., “What would it be like if it were otherwise?” (§418). Implicit in the interlocutory voice’s question in §420, then, is something like the following line of thought:

Isn’t it possible that another person, I mean ‘person’, might lack consciousness, without there being any outward difference in her, I mean ‘her’, behavior? For what has consciousness to do with behavior? Isn’t the former necessarily private and the latter necessarily public? For while I’m certain that I am conscious (and you that you are conscious), I can, at best, only *surmise* that another whom I see is conscious (and another the same of me).⁵⁴ Thus, if it weren’t the case that human beings had consciousness, of course they wouldn’t be ‘unconscious’ in the ‘ordinary sense of the word’, for they’d be *automata*!

Although Wittgenstein’s narrator at first seems to concede that we could, in fact, imagine that others are automata, his response should be read in light of the methodological considerations which I discussed above. Thus, in §420, Wittgenstein’s narrator begins by considering a case in which it would make sense to say that we’re imagining others as automata: “If I imagine it now – alone in my room – I see people with fixed looks (as in a trance) going about their business – the idea is perhaps a little uncanny” (§420). In challenging the interlocutory voice to “try to hang onto this idea in your ordinary intercourse with others” (§420), Wittgenstein’s narrator draws attention to the practical differences between the two settings – i.e., that the latter setting won’t involve the type of activities proper to the former in which one could be said to have imagined that others are automata. Accordingly, Wittgenstein’s narrator anticipates the following difficulties for the interlocutory voice:

⁵⁴ Cf. §246.

Say to yourself, for example: “The children over there are mere automata; all their liveliness is mere automatism.” And you will either find these words becoming quite empty; or you will produce in yourself some kind of uncanny feeling, or something of the sort (§420a).

The expectation isn’t that the interlocutory voice will find his imaginative powers fail where before they were quite successful. Rather, the expectation is that the interlocutory voice realize that he wasn’t doing what he thought he was doing in the former, viz., entertaining a counterfactual.

Thus, the moral of §420 isn’t that we can’t imagine that others are automata, but that, in taking for granted what it means to imagine that others are automata we create the illusion of a substantial difference between creatures with and without consciousness. Moreover, §420 illustrates that what it means to imagine something is hardly unambiguous, either, a point Wittgenstein makes in §397 concerning the extent to which the sense of an expression depends on its imaginability:

397. Instead of “imaginability”, one can also say here: representability in a particular medium of representation. And such a representation may indeed safely point a way to a further use of a sentence. On the other hand, a picture may obtrude itself upon us and be of no use at all.

Wittgenstein’s criticism of the idea that third-personal uses of the concept of consciousness involve indirect, inferential judgments shouldn’t be taken to suggest that he considered such judgements to be direct, or non-inferential, instead – viz., that behavioral criteria provide direct evidence for whether another is consciousness. Rather, Wittgenstein’s criticisms are directed, more broadly, against the idea that third-personal uses of the concept of consciousness involve any appeal to criteria at all. In the next section, I consider what this implies for form of our knowledge in other minds – or, as Wittgenstein puts it in §422: “What do I believe in when I believe that man has a soul?”

4.

Jonadas Techio takes §420 to illustrate the inadequacy of behavioral criteria in determining whether another has a mind, or soul.⁵⁵ Accordingly, the possibility of imagining that others are automata, and furthermore *seeing* others as automata, is taken by Techio to suggest that our seeing and/or treating others *as* human beings depends on more than merely the satisfaction of behavioral criteria. This point is reflected, according to Techio, in Wittgenstein's notion of an "attitude towards a soul" and in Stanley Cavell's distinction between knowledge and acknowledgment in his treatment of the problem of other minds.⁵⁶ On these connections, Techio writes:

The emphasis [...] on our attitudes or reactions (as opposed to our opinions or beliefs), brings to the fore a central aspect of Stanley Cavell's thinking about the "problem of other minds" – namely, that 'the problem' is not a matter of (mere) *knowledge*, but *acknowledgement* [...] These formulations are meant to emphasize that we — that is, each of us — have an active role and an irreducible (although all-too-easily evadable) *responsibility* in adopting a certain attitude in the face of others. This, I take it, is an important first step toward explaining why, even when *all* the behavioral criteria for the ascription of "humanity" are met, one can still avoid adopting that 'attitude towards a soul' of which Wittgenstein speaks, treating those living beings instead as mere automatons.⁵⁷

Moreover, Techio shows considerable interest in §420's concluding analogy, for in his view, it "offers an *explicit* parallel between the experience of seeing aspects in figures and the experience of seeing aspects in living beings (i.e., seeing them *as* automatons / *as* humans)."⁵⁸

⁵⁵ Techio (2013), p. 70.

⁵⁶ Ibid., p. 72.

⁵⁷ Ibid.

⁵⁸ Techio (2013), p. 70.

On this parallel, Techio invokes Stephen Mulhall's reading of Wittgenstein's distinction between the "continuous seeing of an aspect" and the "dawning of an aspect" in *PPF* §118, which he summarizes as follows:

Continuous aspect perception is a "further species of our 'regarding-as' response to pictures — one might say it is our *default response* to them; the experience of aspect-*dawning*, on the contrary, is an exception which proves the rule."⁵⁹

Applying this distinction to §420, Techio writes:

Clearly, we (that is, most of us, most of the time) do not (ordinarily) *take* that we know as human beings *for* human beings, as it would happen in an experience of aspect-dawning. (Human beings are not, *in this sense*, analogous to ambiguous pictures.) Yet, as *PI* [§420] illustrates, in very special circumstances we can stop (avoid, fail) to see human beings as such, and this would be analogous to the (similarly uncanny) experience of making familiar words lose their meanings after much repetition. What this (exceptional) possibility of aspect-change shows, therefore, is that we *continuously see* human beings *as* human, and *in this sense* one can say (as Wittgenstein did) that "[s]eeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another."⁶⁰

Techio's conclusion that we "*continuously see* human beings as human"⁶¹ shouldn't be taken to suggest that he understands there to be a determinate aspect, under which, human beings are continuously seen — as in the case of seeing an ambiguous figure under one or another of its aspects. His qualification that seeing others as automata is more closely analogous to the sort of experience in which "familiar words lose their meaning after much repetition" suggests, instead, that he understands the experience to consist in a disruption to the familiar way in which we're

⁵⁹ Ibid., p. 75.

⁶⁰ Ibid., p. 76.

⁶¹ Ibid.

otherwise oriented toward others, our attitudes towards them – not only in terms of how we see them, but including, also, the range of our diverse responses to them. Borrowing Cavell’s expression, Techio refers to such disruptions as “discrete occurrences of soul-blindness,”⁶² in which, one “avoid[s] adopting that ‘attitude towards a soul’ [...] treating those living beings instead as mere automatons.”⁶³ The moral that Techio draws from §420, then, is that *we* must ultimately maintain that attitude towards a soul, failing which, the possibility of seeing others as automata depends.

Although I share Techio’s sympathies for Cavell’s treatment of the problem of other minds, I want to resist his claim that the inadequacy of behavioral criteria is illustrated by §420 in the experience of seeing others as automata. Towards this goal, I begin by arguing that Techio’s use of the notion of “continuous aspect perception” in the conclusion he draws from §420 that we “*continuously see human beings as human*”⁶⁴ involves a misreading of Wittgenstein’s remarks on aspect perception – the *locus classicus* of which is Chapter xi of *Philosophy of Psychology—A Fragment*, formerly known as Part II of the *Investigations*. I then consider Techio’s suggestion that the experience of seeing others as automata constitutes soul-blindness, or a disruption in one’s attitude towards them, i.e., that one’s attitude towards them no longer takes the form of an attitude towards a soul. Against this, I argue that the experience of seeing others as automata cannot be taken to imply any definite claims about how one is otherwise oriented toward others, or what it means to *see* others as ensouled.

The notion of “continuous aspect perception” is, first of all, an artefact of the secondary

⁶² Ibid., p. 78.

⁶³ Ibid., p. 72; my emphasis.

⁶⁴ Ibid.

literature, for nowhere does the expression appear in Wittgenstein's writings.⁶⁵ Rather, it's derived from what Wittgenstein calls the "continuous seeing" of an aspect in PPF 118, which he distinguishes from the "lighting up" of an aspect immediately after introducing his figure of the duck-rabbit:

118. [...] And I must distinguish between the 'continuous seeing' of an aspect and an aspect's 'lighting up'.

The picture might have been shown me, without my ever seeing in it anything but a rabbit.
[...]

120. I may, then, have seen the duck-rabbit simply as a picture-rabbit from the first. That is to say, if asked "What's that?" or "what do you see there?", I would have replied: "A picture-rabbit." If I had further been asked what that was, I would have explained by pointing to all sorts of pictures of rabbits, would perhaps have pointed to real rabbits, talked about their kind of life, or given an imitation of them.

121. I would not have answered the question "what do you see here?" by saying: "now I see it as a picture-rabbit." I would simply have described my perception, just as if I had said "I see a red circle over there".

Nevertheless, someone else could have said of me: "He sees the figure as a picture-rabbit." At least two notable features of the "continuous seeing" of an aspect emerge from the above remarks. First, as Avner Baz observes⁶⁶, its grammar mirrors that of the first of the two uses of the word "see" that Wittgenstein distinguishes at the beginning of PPF, xi:

111. Two uses of the word "see".

The one: What do you see there?" – "I see *this*" (and then a description, a drawing, a copy).

The other: "I see a likeness in these two faces" – let the man to whom I tell this be seeing the

⁶⁵ Baz (2019), p. 46

⁶⁶ Baz (2019), p. 49.

faces as clearly as I do myself.

What is important is the categorial difference between the two ‘objects’ of sight.

112. The one man might make an accurate drawing of the two faces and the other notice in the drawing the likeness which the former did not see.

113. I observe a face, and then suddenly notice its likeness to another. I *see* that it has not changed; and yet I see it differently. I call this experience “noticing an aspect”.

Cases involving the “continuous seeing” of an aspect therefore take the form of a report – given via a description, a drawing, a copy, etc. – and typically in response to a question such as, “What do you see?” Second, in cases involving ambiguous figures – e.g. the duck-rabbit, Necker cube, etc. – the notion is used only in the third-person, i.e., to say of another person under what aspect they see an object (PPF 121).⁶⁷

Taken together, these features serve to limit the notion of the “continuous seeing” of an aspect to quite specific (types of) cases. On this point, Baz argues:

[The] ‘continuous seeing’ of an aspect as here used by Wittgenstein is only applicable to ambiguous figures—or anyway to things we know may be seen in several more or less determinate ways. It does not apply, for example, to the case of being struck all of a sudden by the similarity of one face to another—where one comes to see that other face in the face one is looking at [...] Here the dawning aspect has not replaced some other aspect under which that face had been seen up until the dawning of the similarity. We were seeing the face all right, but under no particular aspect. And if one wanted to insist that we were seeing the face continuously *as a face*, going against Wittgenstein’s warnings that it makes no sense to talk of seeing something *as* what we *know* it to be, it should then be noted that *that* ‘aspect’ wasn’t eclipsed by the dawning similarity—as the duck is eclipsed by the rabbit, and vice versa—which means that the other alleged ‘aspect’ is

⁶⁷ Ibid., 51.

not grammatically, or ontologically, on a par with the dawning aspect.⁶⁸

On Baz's reading, the notion of the "continuous seeing" of an aspect occupies a rather marginal role in Wittgenstein's remarks on aspect perception, which he distinguishes merely to avoid confusion before ultimately returning his focus to those cases involving the "lighting up" or "dawning" of an aspect.⁶⁹ Baz is therefore critical of those commentators who regard these notions as distinct forms of aspect perception, especially those who have gone even farther in suggesting that "the experience of aspect-dawning should be understood against the background of that continuous perceptual relation to things that may aptly be called 'continuous aspect perception'."⁷⁰ Against this, Baz claims that "Wittgensteinian aspects can *only dawn*,"⁷¹ so that Wittgenstein's stated aim in PPF 115 is really that of "elucidating the concept of *noticing an aspect* and *its* place 'among the concepts of experience'."⁷²

Accordingly, Baz takes Wittgenstein's remarks on aspect perception to be principally concerned with those cases of perceptual experience involving the second use of the word "see" (PPF 111), in which the aspect that "lights up" or "dawns" cannot be given independently of the experience itself. Or, as Baz puts it, "[Wittgensteinian] aspects contrast with what is objectively there to be seen, where what is objectively there to be seen may be determined, and *known* to be there, from a third-person perspective, and independently of any(one's) particular perceptual *experience* of it."⁷³ This doesn't imply that Wittgensteinian aspects are private, ineffable properties of one's subjective experience (cf. PPF 132-4). The "categorical difference between

⁶⁸ Ibid., p. 52.

⁶⁹ Ibid., pp. 48-9.

⁷⁰ Ibid., p. 44.

⁷¹ Ibid., p. 10.

⁷² Ibid., p. 11.

⁷³ Ibid., p. 11; my emphasis.

the two ‘objects’ of sight” that Wittgenstein considers in *PPF* 111 is not an ontological, but a grammatical one.

To make this last point clear, I want to return to the example Wittgenstein considers in *PPF* 111-2. I noted above that cases involving the first use of the word “see” paradigmatically take the form of a report, such that one can *inform* another person of what one sees in response to such a question as, “What do you see *there*?” Baz adds to this that “unless [the person who asks this] is testing our eyesight or linguistic competence, she is asking because she cannot, for some more or less contingent reason, see for herself.”⁷⁴ On the contrary, the point of giving expression to the dawning of an aspect is not to inform another person of what one *sees* (according to the second use of the word “see”), where the other isn’t in a position to see it. Rather, as the example of *PPF* 111-2 illustrates, two, similarly-situated individuals might see (according to the first use of the word “see”) the same two faces, while one notices in it a likeness which the other fails to *see* (according to the second use of the word “see”). The former might nonetheless give expression to her experience by inviting the other to notice the likeness that she had hitherto failed to *see* (according to the second use of the word “see”). Thus, as Baz notes, “Wittgensteinian aspects are importantly characterized by the possibility that a fully competent speaker (and perceiver) may fail to see (or otherwise perceive) them even though she sees (first sense) as well as anyone else the objects in which they are seen, and by the particular sense it makes to *call upon* such a person to see them.”⁷⁵

It is in this sense, then, that aspects – unlike the third-personal, objective features of the world with which they’re contrasted – are, as Wittgenstein puts it, “subject to the will”:

256. Seeing an aspect and imagining are subject to the will. There is such an order as “Imagine *this!*”,

⁷⁴ *Ibid.*, p. 12.

⁷⁵ *Ibid.*, pp. 12-3

and also, “Now see the figure like *this!*”; but not “Now see this leaf green!”.

Thus, what the grammatical difference between the two uses of the word “see” suggests is that the point of seeing aspects is not to inform another of what one sees (according to the first use of the word “see”) but rather, Baz argues, “to *invite* the other to see something in or about the object – something that strikes us as there to be seen, even though we normally acknowledge that there is no way for us to establish its presence.”⁷⁶ In other words, to insist on treating aspects as if their presence were objectively confirmable is, precisely, to *miss the point* of seeing aspects.

Applying the above considerations to Techio’s reading of §420, the following problems emerge. First, his conclusion presumes that seeing human beings as automata is tantamount, or reducible to a failure to see human beings *as* human beings. If we grant that seeing other human beings as automata is, in some sense, an extraordinary or anomalous experience, it doesn’t therefore follow that we ordinarily see other human beings *as* human beings. For the aspect that dawns, or “lights up”, in the case of seeing others as automata doesn’t replace any particular way in which we might have otherwise been seeing others. Even assuming there were an “aspect” under which we were seeing others up until the moment we began seeing them as automata, that “aspect” wouldn’t be, as Baz puts it, “grammatically, or ontologically on a par with the dawning aspect.”⁷⁷ For it cannot be said that we “*continuously see* human beings as human”⁷⁸, if “see” is used here in the same sense in which we say that we “see living human beings as automata” (§420).

Despite qualifying that he doesn’t take there to be a determinate aspect under which human beings are continuously seen as human beings, Techio’s suggestion that the experience of

⁷⁶ Baz (2011), p. 711. See also: Baz (2000).

⁷⁷ Baz, p. 52.

⁷⁸ Techio (2013), p. 76.

seeing others as automata constitutes an anomalous disruption in our attitudes towards others remains problematic. Indeed, this line of thought is in entirely the wrong direction, for we would, in effect, produce in ourselves that “uncanny feeling” (§420) about which Wittgenstein’s narrator warned. That is, it would lead us anxiously to wonder at our condition vis-a-vis others; e.g. what is it about others, my relation to others, *me*, that accounts for the possibility of seeing them otherwise than as I normally (or should?) see them? (cf. “The transition from obvious nonsense to something which is unobvious nonsense” (§524).)

The mistake, then, is to interpret the uncanniness of the experience of seeing others as automata to imply anything about how we ordinarily see others. In §596, Wittgenstein considers a similar mistake underlying a particular misconception regarding the relation between the feelings of familiarity and unfamiliarity, i.e., the idea that the latter constitutes a disruption of the former:

[T]here are feelings of strangeness: I stop short, look at the object or man questioningly or suspiciously, and say “I find it all strange”. – But the existence of this feeling of strangeness does not give us a reason for saying that every object which we know well and which does not seem strange to us gives us a feeling of familiarity – It is as if we thought that the space once filled by the feeling of strangeness must surely be filled by *something*. The space for these kinds of atmosphere is there, and if one of them is not filling it, then another is (§596).

Thus, although Techio doesn’t take there to be a determinate aspect under which human beings are continuously seen, he nonetheless makes the same type of mistake in attempting to account for the feeling of uncanniness produced by the experience of seeing others as automata by treating the situation as if there were a specifiable *something* for which that feeling is a replacement – viz., an “attitude towards a soul.” Indeed, that Techio takes there to be role along these lines for Wittgenstein’s notion of an “attitude towards a soul” suggests a serious

misunderstanding.⁷⁹

Finally, I want to address this last point by looking briefly at the set of remarks in *PPF* from which Wittgenstein's notion of an "attitude towards a soul" is taken:

19. "I believe that he is suffering." — Do I also *believe* that he isn't an automaton?

Only reluctantly could I use the word in both contexts.

(Or is it like *this*: I believe that he is suffering, but am certain that he is not an automaton? Nonsense!)

20. Suppose I say of a friend: "He isn't an automaton." — What information is conveyed by this, and to whom would it be information? To a *human being* who meets him in ordinary circumstances? What information *could* it give him? (At the very most, that this man always behaves like a human being, and not occasionally like a machine.)

21. "I believe that he is not an automaton", just like that, so far makes no sense.

22. My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul.

As the concluding parenthetical of §20 intimates, the claim "He isn't an automaton" is uninformative because it does not offer any meaningful contrast with what the friend, as a matter of fact, *is* – as if there were *something* that the friend possessed, something peculiar to human beings and peculiarly absent from automata (viz. consciousness, a soul), in virtue of which, saying "He isn't an automaton" could be taken to mean "He is a human being". Moreover, Wittgenstein pursues a similar strategy to that of §§412-27 in suggesting that, grammatically speaking, the point of such remarks as "He isn't an automaton" isn't to inform others. That isn't

⁷⁹ There is an extensive literature on this topic, including the following two pieces from a recently published anthology, *Moral Foundations of Philosophy of Mind*: Cerbone (2019) and Dain (2019). The following is a sample of some of the more well-known contributions to that literature: Cook (1969), Cavell (1979) Winch (1980-1), and Ter Hark (1991).

to say that such a remark couldn't be used significantly, but only that it shouldn't be understood as specifying what, as a matter of fact, he (or any human being) isn't. The upshot of PPF §§19-22 parallels §§416-20; namely, that third-personal uses of the concept of consciousness (or the soul) don't involve indirect, or inferential judgements about what a human being, as a matter of fact, is, for the form of our knowledge of other minds isn't one of matter-of-factual belief, or opinion.

In describing the form of our knowledge of other minds as an "attitude towards a soul", instead, Wittgenstein shouldn't be understood as suggesting a technical distinction between an attitude and an opinion, according to which the former is, so to speak, an epistemically stronger form of the latter. Nor does Wittgenstein use "attitude" in the sense of a disposition or outlook, as if it were a particular sort of inclination or way of looking at things. Indeed, that Wittgenstein doesn't intend "attitude" in a cognitive sense is evident from his use of the term in §310 of the *Investigations*:

310. I tell someone I'm in pain. His attitude to me will then be that of belief, disbelief, suspicion, and so on.

Let's suppose he says, "It's not so bad". – Doesn't that prove that he believes in something behind my utterance of pain? — His attitude is proof of his attitude. Imagine not merely the words "I'm in pain", but also the reply "It's not so bad", replaced by instinctive noises and gestures.

In suggesting that doubts regarding another's expression of pain will correspond to differences in one's attitude towards the other shouldn't be taken to imply that an "attitude towards a soul" will be (at least temporarily) suspended in the face of such doubts. As the subsequent paragraph suggests, one's attitude towards another is characteristically unreflective or instinctual, such that the changes are understood to describe differences at the level of "fine shades of behavior" (PPF

§210). Ultimately, however, such alterations leave unaffected the condition of one's attitude towards a soul.

Thus, although the experience of seeing another human being as an automaton might affect one's attitude towards her – leading it doesn't follow that one's attitude will no longer be an attitude towards a soul. For in such cases, the difference will occur at the level in which attitudes of belief, disbelief, and suspicion are discriminated. To suppose otherwise, as Techio does, would which the concepts of a human being and an automaton are the internal way in which the concepts of a human being and thus overlook the moral of the analogy with which §420 concludes: "Seeing a living human being as an automaton is analogous to seeing one figure as a limiting case or variant of another; the cross-pieces of a window as a swastika, for example" (§420). As the example of seeing a window's cross-pieces as a swastika suggests, seeing the windows cross-pieces as a swastika doesn't consist in a *failure* to see the window as such. Rather, the possibility for that change of aspect depends on conceiving the window-frame as a limiting case of a swastika. Likewise, the possibility of seeing a human being as an automaton depends on conceiving human beings as limiting cases of automata – as, for example, Descartes did.

The mistake, according to Wittgenstein, is to assume that a particular mode of representation can be justified or rejected in terms of whether it more or less accurately describes the facts. Indeed, one of most profound themes of the *Investigations* is Wittgenstein's recognition of the sources of philosophical problems in those instances in which we're led to believe that "our way of speaking does not describe the facts as they really are" (§402). As Cora Diamond explains:

Wittgenstein describes us as "tempted to say that our way of speaking does not describe

the facts as they really are.” And this is not because he thinks that it *does* describe the facts as they are, but rather because he takes our mode of speaking not to describe any facts at all. More strongly: if we had a different way of speaking, we should not be getting something in reality wrong that we are now getting right, nor should we be getting something in reality right that we are now getting wrong. There are no metaphysical facts to make our way of speaking right or wrong; there is nothing *out there* to make the necessities we have built into language correct or incorrect.⁸⁰

This recognition is intimately related to Wittgenstein’s conception of philosophy as a clarificatory activity – his insistence that “philosophy must interfere in any way with the actual use of language, so [that] it can in the end only describe it” (§124; my emphasis). That philosophy “leaves everything as it is” (§124) doesn’t imply an appeal to common sense, as it were, but as Diamond explains, the recognition that “philosophy does not put us in a position to justify or criticize what we do by showing that it meets or fails to meet requirements we lay down in our philosophizing.”⁸¹

⁸⁰ Diamond (1991), p. 15.

⁸¹ Ibid., p. 22.

Bibliography

- Baz, Avner (2000). "What's the Point of Seeing Aspects?" *Philosophical Investigations*, Vol.23(2). 97-121.
- (2011). "Aspect Perception and Philosophical Difficulty", *The Oxford Handbook of Wittgenstein*, edited by Marie McGinn and Oskari Kuusela, Oxford University Press. 697-713.
- (2019). *Wittgenstein on Aspect Perception* (forthcoming).
- Beckermann, Ansgar (2004). "Wittgenstein, Wittgensteinianism, and the Contemporary Philosophy of Mind: Continuities and Changes", in *Wittgenstein Today*, pp. 275-95.
- Cavell, Stanley (1962). "The Availability of Wittgenstein's Later Philosophy," *The Philosophical Review*, Vol.71(1).
- (1979). *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*, Oxford University Press.
- Cerbone, David (2019). "All Souls: Wittgenstein and 'eine Einstellung zur Seele'", in *Moral Foundations of Philosophy of Mind*, edited by Joel Backström Joel, Hannes Nykänen, Niklas Toivakeinen, and Thomas Wallgren, Palgrave Macmillan, pp. 79-95.
- Chalmers, David (1995). "Facing Up to the Hard Problem of Consciousness", in *Journal of Consciousness Studies*, Vol.2, pp. 200-19.
- (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.
- (2002). "Consciousness and Its Place in Nature", in *Philosophy of Mind: Classical and Contemporary Readings* (2002), edited by David J. Chalmers, Oxford University

- Conant, James (2005). "Stanley Cavell's Wittgenstein," *The Harvard Review of Philosophy* Vol.13(1).
- Cook, John (1969). "Human Beings", *Studies in the Philosophy of Wittgenstein*, edited by Peter Winch, London: Routledge and Kegan Paul. 117-51.
- Dain, Edmund (2019). "Attitude Towards a Soul: Wittgenstein, Other Minds, and the Mind", in *Moral Foundations of Philosophy of Mind*, edited by Joel Backström Joel, Hannes Nykänen, Niklas Toivakeinen, and Thomas Wallgren, Palgrave Macmillan, pp. 79-95.
- Dennett, Daniel (1991). *Consciousness Explained*, London: Penguin Books.
- (1995). "The Unimagined Preposterousness of Zombies", Symposium on "Conversations with Zombies", in *Controversies in Science and the Humanities*, Vol.2(4), pp. 322-6.
- (1999). "The Zombie Hunch: Extinction of an Intuition?" *Royal Institute of Philosophy Millennial Lecture*, pp. 27-43.
- (2002). "Quining Qualia", in *Philosophy of Mind: Classical and Contemporary Readings* edited by David J. Chalmers, Oxford University Press, pp. 226-56.
Reprinted from *Consciousness in Contemporary Science*, edited by A. Marcel and E. Bisiach with permission of Oxford University Press.
- (2003). "Who's on First? Heterophenomenology Explained", in *Journal of Consciousness Studies*, Vol.10(9-10), pp. 19-38.
- Diamond, Cora (1991). (1991). *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, MIT Press.
- Hutto, Daniel (1995). "Consciousness Demystified: A Wittgensteinian Critique of Dennett's Project", in *The Monist*, Vol. 78(4), pp. 464-79.

- Schulte, Joachim (2017). "Body and Soul," *The Cambridge Companion to Wittgenstein*, 2nd edition, edited by Hans Sluga and David Stern. Cambridge University Press.
- Stern, David (2004). *Wittgenstein's Philosophical Investigations: An Introduction*. Cambridge University Press.
- (2007). "The uses of Wittgenstein's beetle: Philosophical Investigations §293 and its interpreters." In *Wittgenstein and his Interpreters*, edited by Guy Kahane, Edward Kanterian, and Oskari Kuusela, pp. 248-268. Blackwell.
- (2017). "Wittgenstein's Texts and Style." In *A Companion to Wittgenstein* ed. Hans-Johann Glock and John Hyman, 41-55. Blackwell.
- Techio, Jônadas (2013). "Seeing Souls: Wittgenstein and Cavell on the 'Problem of Other Minds.'" *Conversations: The Journal of Cavellian Studies*. Issue 1.
- Ter Hark, Michel (1991). "The Development of Wittgenstein's Views about the Other Mind's Problem", *Synthese*, Vol. 87(2), Wittgenstein, Part II. 227-53.
- Winch, Peter (1980-1). "Eine Einstellung zur Steele", *Proceedings of the Aristotelian Society*, New Series, Vol. 81. pp. 1-15.